AD-A041 759 THRESHOLD TECHNOLOGY INC DELRAN N J F/G 9/2
ALPHA/NUMERIC EXTRACTION TECHNIQUE PHASE II.(U)
MAY 77 P B SCOTT F30602-76-C-0309
UNCLASSIFIED RADC-TR-77-184 NL

1 OF 1
AD
A041759

END
DATE
FILMED
8-77

RADC-TR-77-184
Final Technical Report
May 1977

ALPHA/NUMERIC EXTRACTION TECHNIQUE PHASE II

Threshold Technology, Inc.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.

This report has been reviewed and is approved for publication.

APPROVED: Richard S. Vonusa

RICHARD S. VONUSA
Project Engineer

APPROVED: H Davis

HOWARD DAVIS
Technical Director
Intelligence and Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

19

1. REPORT NUMBER: RADC-TR-77-184 (18)

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): ALPHA/NUMERIC EXTRACTION TECHNIQUE PHASE II (6)

5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report (9)

6. PERFORMING ORG. REPORT NUMBER: N/A

7. AUTHOR(s): Phillips B. Scott (10)

8. CONTRACT OR GRANT NUMBER(s): F30602-76-C-0309 (15) new

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Threshold Technology, Inc.
Delran NJ 08075

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
64701B
43020309 (16) (17) φ3

11. CONTROLLING OFFICE NAME AND ADDRESS:
Rome Air Development Center (IRAP)
Griffiss AFB NY 13441

12. REPORT DATE: May 1977 (11)

13. NUMBER OF PAGES: 27

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same

(12) 35 p.

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Richard Vonusa (IRAP)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Recognition
Pattern Recognition
Acoustic Phonetics

2-part (1) (2)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes a ~~two-part~~ program to interface an existing word recognition system (WRS) to a bathymetric digitizer and to develope a new word recognition system for a large vocabulary. The ~~interfacing of a WRS with a bathymetric digitizer~~ first program provides an improved means for entering bathymetric readings from smooth sheets to punched cards. The WRS allows ~~the~~ a cartographer ~~to~~ to simultaneously obtain X-Y coordinate locations and provide voice data entry of depth reading for each coordinate location. With the operator's hands free to concentrate on the X-Y position sensing device (cursor), the

cont on (p1473B)

DD FORM 1 JAN 73 1473A EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390 798

mt

(cont fr p1473A)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

she ~~operator~~ can speak the depth numbers. If they are correctly recognized, he ~~or~~ ~~she~~ can enter them onto punched cards without losing sight of the smooth sheet. The cartographic WRS which was interfaced to the digitizer ~~was developed under contract F30602-75-C-0238 by Threshold Technology Inc. (TTI).~~ is described in AD-B008303L The complete system has been installed at the Defense Mapping Agency Hydrographic Center ~~(DMAHC)~~ at Suitland, MD. The WRS at DMAHC has a vocabulary of the ~~ten~~ digits plus four control words. It can store reference data for five speakers. A special provision has been made to allow correction of previously entered depth data. Any of the last five depth entries can be corrected at any time. Cards containing X, Y and Z data are punched automatically.

The large vocabulary WRS developed for this program is intended to ~~be~~ will used to explore other cartographic applications. This system is based upon the VIP-100, a WRS manufactured commercially by TTI. This system has the system with the capability of recognizing up to 200 separate words in a syntactic structure. The structure allows up to 30 words in any node of the structure. Up to 30 nodes may be included in a sentence structure. Node structures can be changed by simple Teletype commands and node plans can be stored on floppy disks by means of a disk drive which is a part of the system. The minicomputer included in the system is a Data General Nova 2 with 16K of core memory. The WRS is speaker dependent with reference data for one speaker residence in core at any time. Speaker data can also be stored on a floppy disk.

ACCESSION for		
NTIS	White Section	☑
DDC	Buff Section	☐
UNANNOUNCED		☐
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL. and/or SPECIAL	
A		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

EVALUATION

This program has resulted in the fabrication of an Advanced Development Model (ADM) Limited Vocabulary Word Recognition System. The ADM word recognition data entry system allows an operator to enter into a computer bathymetric (ocean depth measurements) data by voice rather than by the conventional manual keyboard method. A distinct advantage of voice input is that the operator has both of his hands free and can devote his full attention to accurate positioning of the cursor.

Laboratory experiments utilizing voice data entry equipment have proven that an individual unfamiliar with the voice recognition technology is capable on the average of inputting 600 four-digit depth soundings per hour.

A briefing of the voice technology was presented by RADC to Defense Mapping Agency Hydrographic Center (DMAHC) personnel. Voice recognition hardware was also demonstrated in order to verify a 600 four-digit throughput rate. As a result of the demonstration, DMA decided that the voice technology would significantly improve the production rate for bathymetric digitization for automated chart development. DMAHC recommended that the technology be implemented on a future program development called, "Bathymetric Data Reduction System" (BDRS).

The ADM delivered under this program has been installed at DMAHC in Suitland, Maryland for a six month test in a production environment. The purpose of the test is to familiarize DMAHC operators with voice digitization and discover and correct any deficiencies in the voice technology so that the BDRS voice recognition module is optimized for a production run environment.

This program represents the first time that voice recognition technology has successfully been implemented in an actual operational system. Much has been learned from developing this system and the task of integrating voice technology into more complex DMA applications such as Digital Radar Landmass Simulator (DRLMS) and production of Flight Information Publications (FLIPS) will be made much easier.

Richard S. Vonusa
RICHARD S. VONUSA
Project Engineer

Section I

INTRODUCTION

An advanced development model limited-vocabulary word recognition system (WRS) was constructed by TTI for cartographic use under contract F30602-75-C-0238.[1] This equipment was originally intended to be used with a bathymetric digitizing system located at RADC to input bathymetric depth readings from smooth sheets. During the present effort the WRS was moved to the Defense Mapping Agency Hydrographic Center (DMAHC) in Suitland, Maryland. At the DMAHC the WRS was interfaced with a Computer Equipment Company Digi-Grid system model DG10G digitizing table. Extensive redesign of interfacing hardware and software was necessary to allow the WRS to be used with the digitizing table. The resulting combination performs a similar function to the RADC system except that the output medium of the DMAHC system is punched cards rather than a serial record of X, Y, and Z values on magnetic disc or tape as was the case at RADC.

Also developed during this program was an advanced development model of a highly reliable isolated-word, speaker dependent, limited vocabulary word recognition system. This system recognizes up to 200 words arranged in a structured manner. The structure consists of any combination of nodes. Each node can include up to 30 vocabulary words. The system which is based upon the VIP-100 word recognition system includes a Data General Nova 2 computer with 16K of core memory. A 16 character alphanumeric self-scan display shows recognition decisions of the system. A Xebec Flexible Disc System model XFD-100 complete with two drives, controller and Nova interface is included in the system. The flexible disc memory provides a convenient medium for storage of speaker reference data for up to 20 speakers per disc. Also a multiplicity of node plans can be stored on disc.

Section II

DMAHC SPEECH INPUT BATHYMETRIC DIGITIZING SYSTEM

A. Introduction

The Defense Mapping Agency Hydrographic Center (DMAHC) in Suitland, Maryland has the task of digitizing water depth readings from charts covering all parts of the globe. Depth data has been recorded by hand for decades on thousands of charts covering water areas from shallow harbors to deep oceans. Heretofore, the DMAHC has manually accomplished digitizing of depth data together with the X and Y coordinates associated with the depth readings. Two-dimensional coordinate digitizing systems which include a digitizing table, a free cursor and associated electronics have been used. With these systems, each depth reading is inputted by keyboard after the cursor is placed over the depth value on a chart in order to ascertain X and Y coordinates which are correlated with the region of the earth covered by the chart. A bathymetric operator involved in this process must constantly shift his or her attention from the cursor to the keyboard and back. By replacing the keyboard with a voice input system the operator's attention can be fixed on the exact location of the cursor and accurate depth digit entry. The first phase of this effort was concerned with the interfacing of a WRS (built for RADC under contract F30602-75-C-0238) with a digitizing system at DMAHC. The following paragraphs describe the configuration and operation of the interfaced system.

B. System Configuration

The system configuration at DMAHC is shown in the block diagram of Fig. 1. The Nova 1200 minicomputer serves as the central interface with all system components. Speech data is processed by the Nova 1200 in the same manner described in the final report for contract F30602-75-C-0238.[1] Cursor locating coordinates (X and Y data) are received by the computer by the use of an RS232 interface at 2400 baud from the digitizer controller which senses the location of the cursor on the digitizing table. The minicomputer, after recognizing input depth data, consisting of one to four digits, combines this depth data with the X and Y coordinate data from the digitizer and drives an IBM 29 card punch by means of card punch interface in the minicomputer. The two displays are driven by a single channel between the self-scan display module and the computer. Individual system components are described below insofar as modifications or additions are concerned.

1. Digitizer

The digitizer to which the WRS has been interfaced is the Computer Equipment Company (CEC) model DG-10C. In July 1976, CEC was sold to the Altek Company of Rockville, Maryland, also a manufacturer of digitizing tables and equipment. The digitizer is an all electronic two-dimensional coordinate digitizing system which uses a "free cursor" for line and curve tracing. This system was originally connected directly to the IBM 29 card punch. Depth data was inputted by the use of a keyboard. Altek Corp. added an EIA RS232

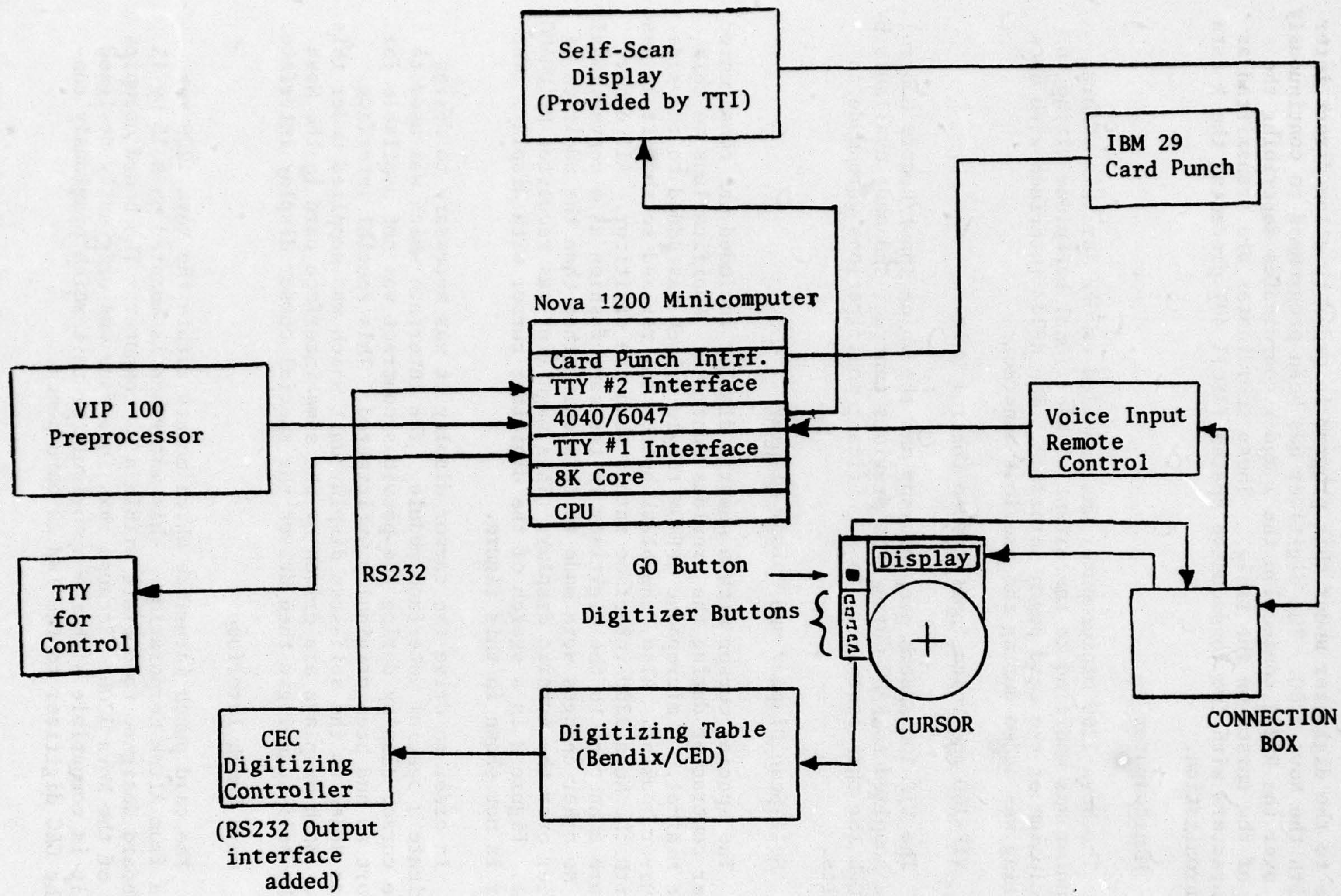

Fig. 1 Block diagram of DMAHC speech digitizing system.

interface to the digitizer under this program in order to allow direct interfacing with the Nova 1200. The digitizer has been programmed to continuously transmit over the RS232 connection the X and Y coordinates describing the location of the cursor on the table. These coordinates are transmitted as ASCII characters with two consecutive spaces (ASCII 40) preceding the X data for synchronization.

2. Minicomputer

The Nova 1200 minicomputer was supplied as GFE for this effort. This computer was modified to the extent of additional backplane wiring to allow inclusion of the card punch interface. The RS232 interface with backplane wiring was added during the previous contract.

3. VIP-100 and Voice Input/Remote Control Unit

The VIP-100 speech preprocessor and the Voice Input/Remote Control Unit were supplied to RADC during the previous contract and made available by RADC as GFE for this contract. No significant modifications were made to these units.

4. Self-Scan Display and Cursor Display

The special cursor with a numeric display included was constructed by another contractor during the previous contract. Modifications to this unit were minimal. A microphone OFF-ON toggle switch was added to the side of the cursor housing. Also, the black button was revised so that it is associated with the Nova 1200 interface rather than the digitizer. All other buttons are connected to the digitizer in the same fashion as a conventional cursor. No other changes were made to this unit other than the addition of a magnifier over the numeric display. This magnifier was requested by DMAHC personnel. Figure 2 is a sketch of the digitizer cursor with display. The magnifier is not shown in this figure.

In order to drive the cursor display it was necessary to design and fabricate a special interface module. The interface which was used to drive the cursor display during the previous contract was not available for this effort as had been originally anticipated. This special interface module was added to the self-scan display unit which was supplied under this contract. Both displays are driven by the same interface card in the Nova 1200. The self-scan logic then drives the special cursor display interface.

5. Card Punch Interface

The card punch interface which mounts within the Nova 1200 was purchased from Altek Corporation. This interface is mounted on a 15 by 15 inch PC board designed to operate within a DG computer. The board occupies slot six of the Nova 1200. Because this interface was originally designed by CEC, it is compatible with the opto-isolator unit which previously connected the CEC digitizer to the IBM 29 card punch.

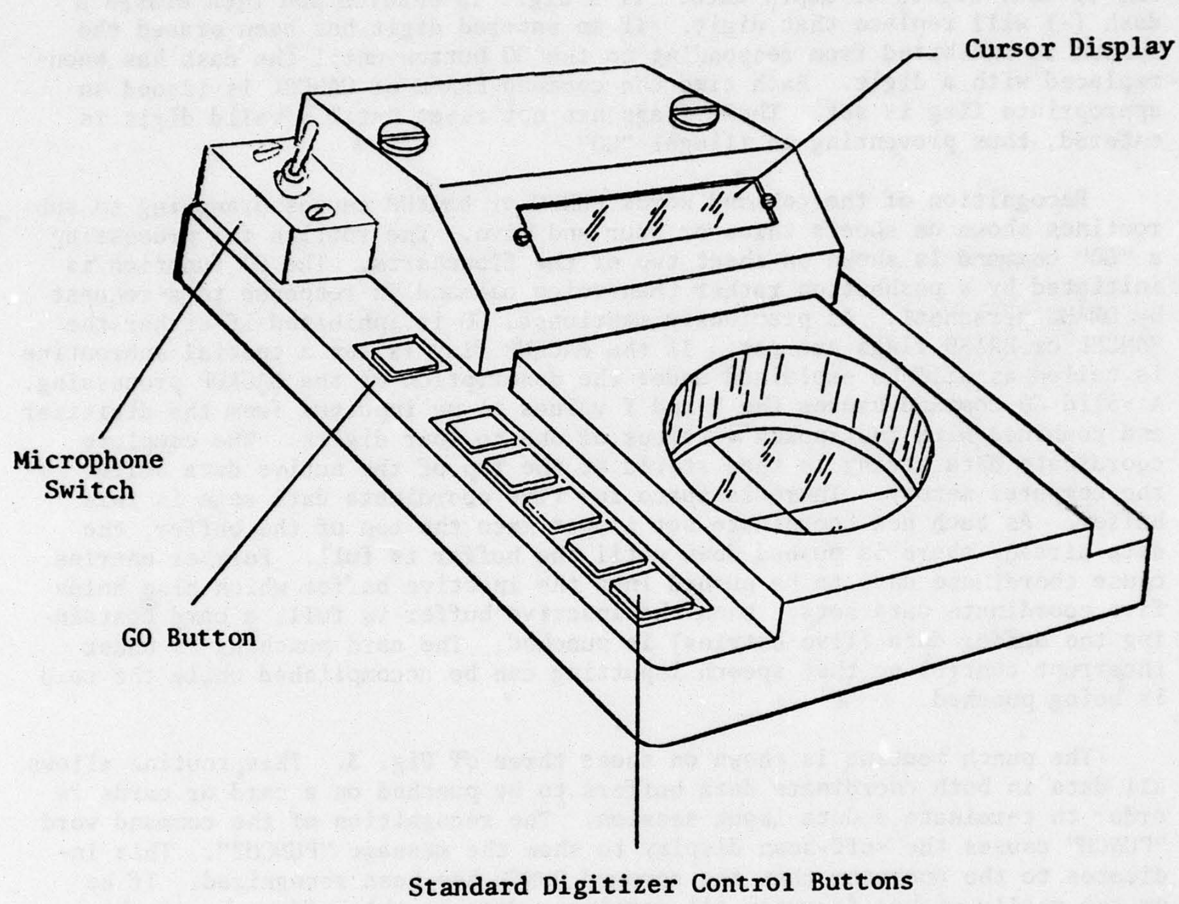

Fig. 2 Modified digitizer cursor with display.

C. Software

Except for the word recognition routine, the DMAHC software is completely new. Flowcharts for this software are shown in Figure 3. Sheet one (of five) illustrates the main control loop of the program. When the system is initialized the self-scan display shows the message "SYSTEM IS READY". At the same time, the cursor display is dark. The system is indeed ready to accept from one to four digits of depth data. If a digit is entered and then erased a dash (-) will replace that digit. If an entered digit has been erased the system is inhibited from responding to the GO button until the dash has been replaced with a digit. Each time the command ERASE or CANCEL is issued an appropriate flag is set. These flags are not reset until a valid digit is entered, thus preventing an illegal "GO".

Recognition of the command words PUNCH or BACKUP causes branching to subroutines shown on sheets three or four and five. The routine for processing a "GO" command is shown on sheet two of the flowcharts. The GO function is initiated by a pushbutton rather than voice command in response to a request by DMAHC personnel. As previously mentioned, GO is inhibited if either the CANCEL or ERASE flags are set. If the BACKUP flag is set a special subroutine is called as will be explained under the description of the BACKUP processing. A valid GO command causes the X and Y values to be inputted from the digitizer and combined with the spoken Z values of one to four digits. The complete coordinate data string is then stored at the top of the active data buffer in the computer memory. There is space for five coordinate data sets in this buffer. As each new coordinate set is put into the top of the buffer, the data already there is pushed down until the buffer is full. Further entries cause coordinate data to be pushed into the inactive buffer which also holds five coordinate data sets. When the inactive buffer is full, a card containing the buffer data (five entries) is punched. The card punching is under interrupt control so that speech inputting can be accomplished while the card is being punched.

The punch routine is shown on sheet three of Fig. 3. This routine allows all data in both coordinate data buffers to be punched on a card or cards in order to terminate a data input session. The recognition of the command word "PUNCH" causes the self-scan display to show the message "PUNCH?". This indicates to the operator that the command PUNCH has been recognized. If he or she really wishes to punch all remaining data in the buffer, he or she must say PUNCH again. The data remaining will then be punched and the system will be completely initialized. Requiring the operator to say "PUNCH" two times obviates the problem which could occur if a misrecognition occurred on another word. After PUNCH has been said once only PUNCH or ERASE will be accepted by the system.

The BACKUP routine shown by flowcharts four and five provides a means for the operator to correct by voice any of the last five depth entries at any time. This routine is called by the command word "BACKUP" which is active after the "GO" button has been pressed and before new depth data has been entered. The first time BACKUP is said a pointer is set to the top of the active data buffer. The depth entry at the top of the buffer is displayed on the cursor. Actually the cursor display will not change at this point

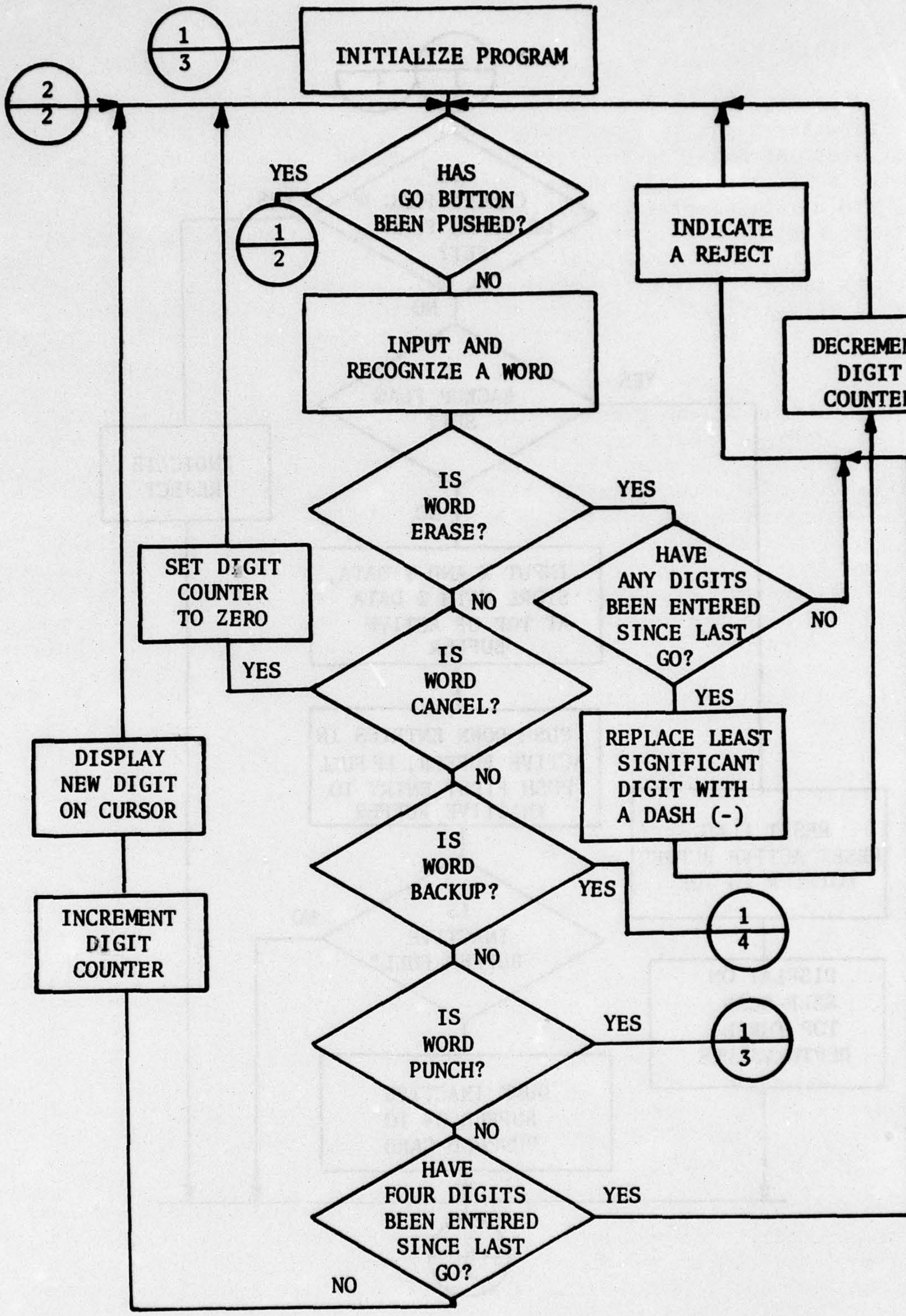

Fig. 3 Flowchart of DMAHC program. (Sheet 1 of 5)

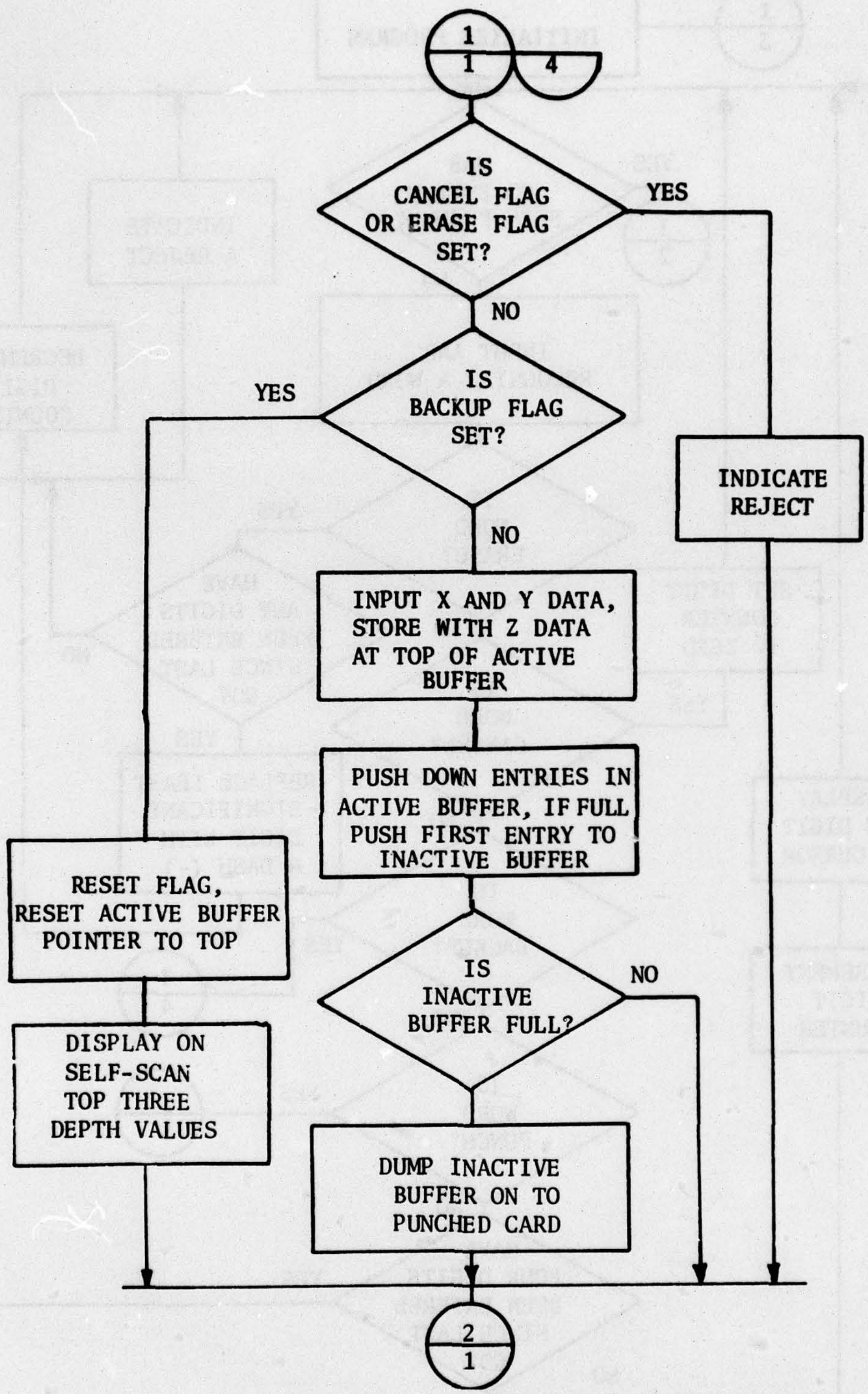

Fig. 3 Flowchart of DMAHC program. (Sheet 2 of 5)

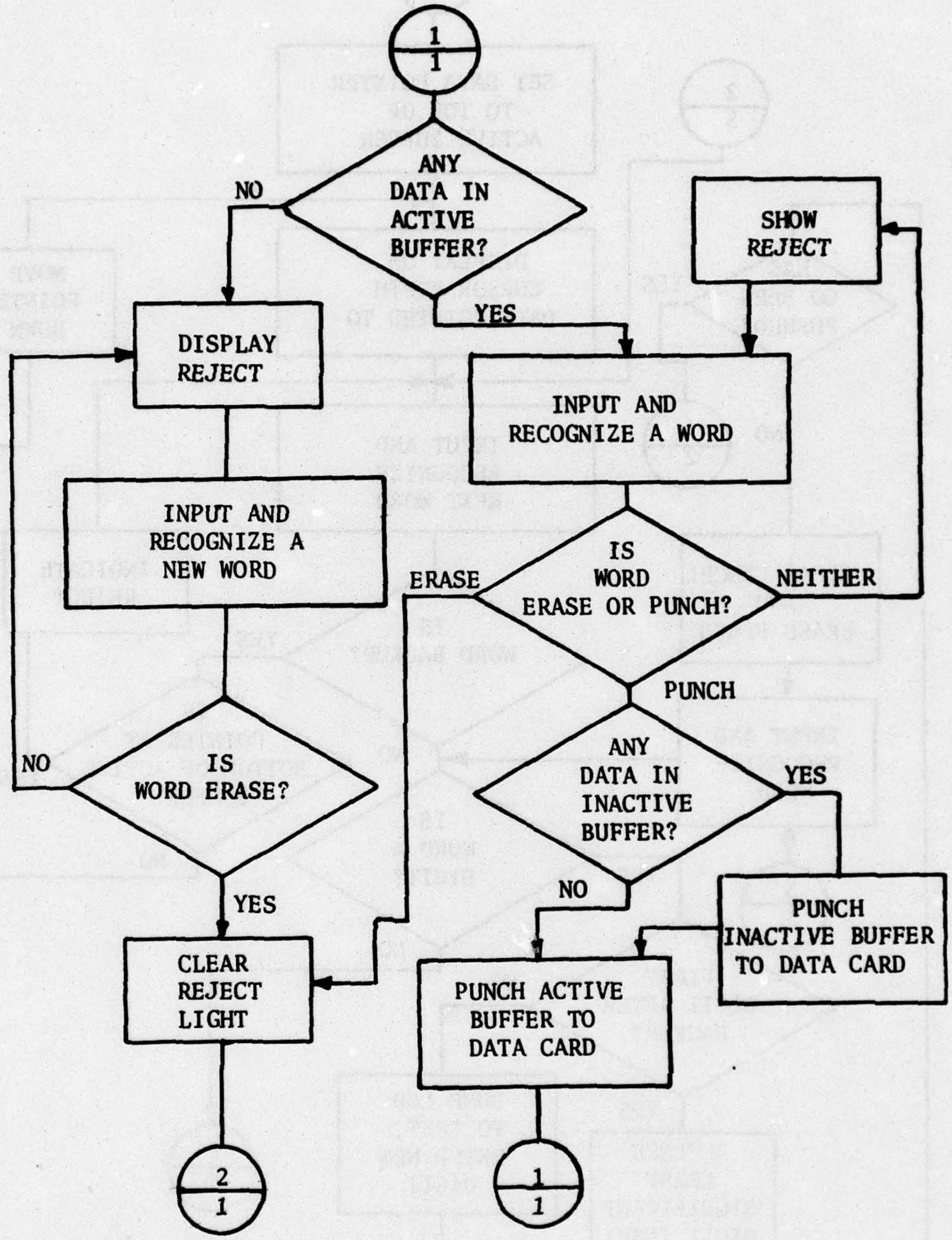

Fig. 3 Flowchart of DMAHC program. (Sheet 3 of 5)

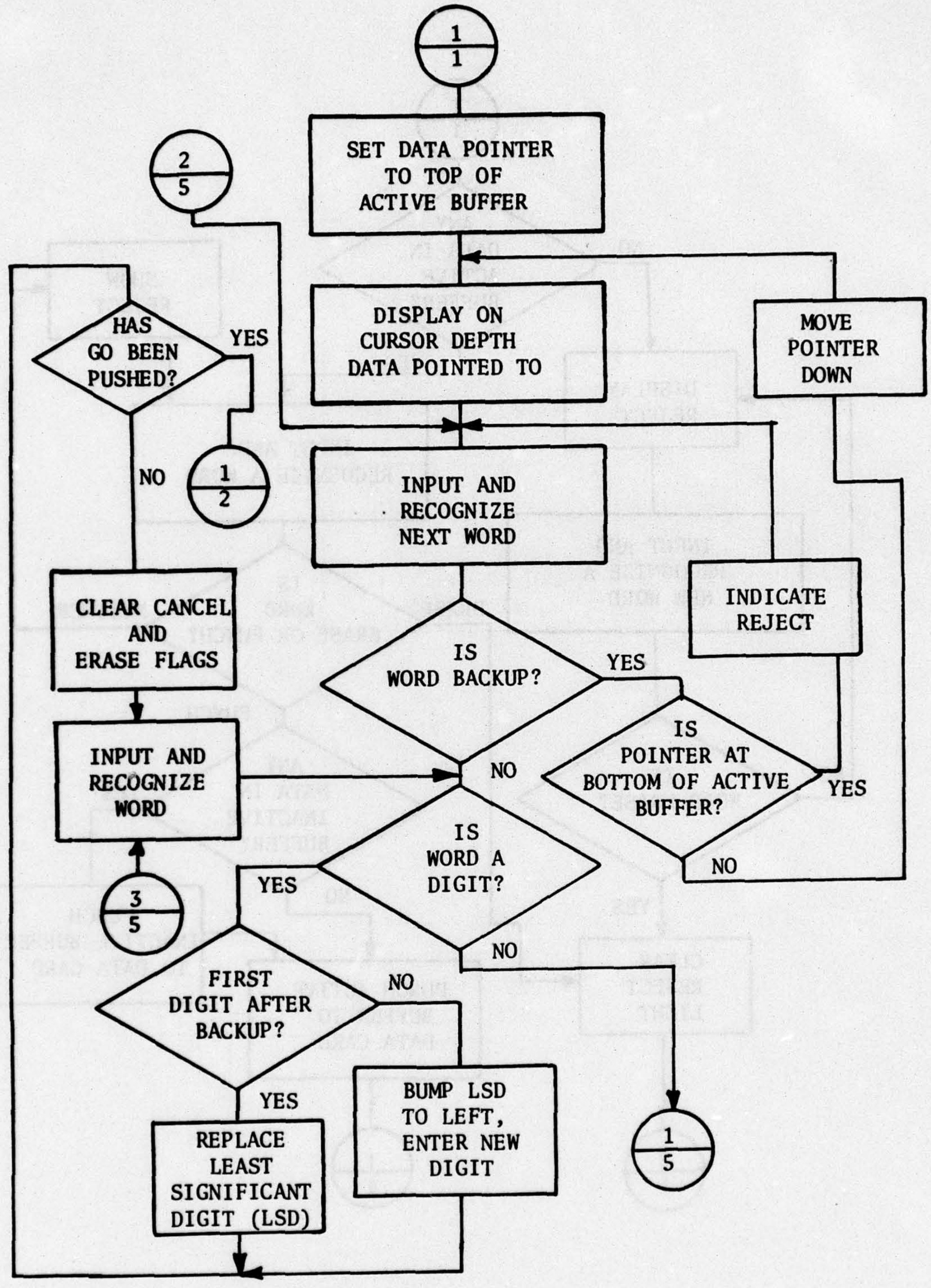

Fig. 3 Flowchart of DMAHC program. (Sheet 4 of 5)

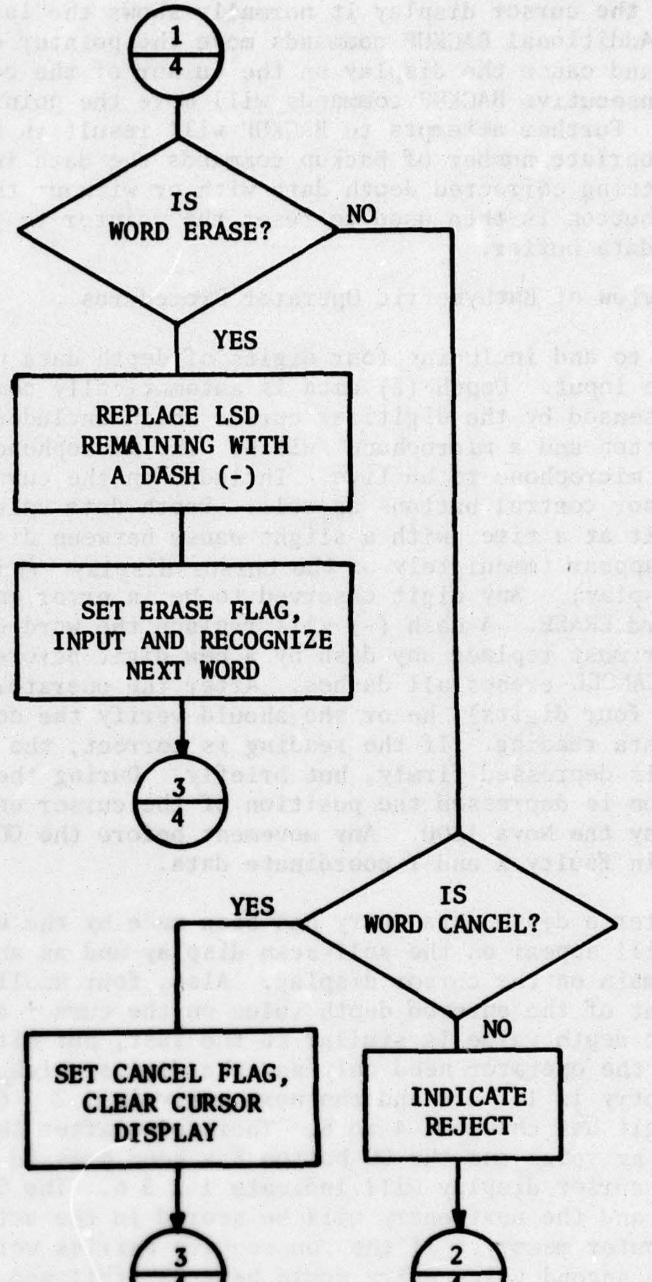

Fig. 3 Flowchart of DMAHC program. (Sheet 5 of 5)

because the cursor display it normally shows the last depth input before a "GO". Additional BACKUP commands move the pointer down through the active buffer and cause the display on the cursor of the corresponding depth data. Five consecutive BACKUP commands will move the pointer to the bottom of the buffer. Further attempts to BACKUP will result in reject indications. After an appropriate number of backup commands the data in error can be corrected by inputting corrected depth data with or without the aid of ERASE or CANCEL. The GO button is then used to reset the pointer to just above the top of the active data buffer.

D. Review of Bathymetric Operator Procedures

Up to and including four digits of depth data per sounding can be entered by voice input. Depth (Z) data is automatically combined with X and Y coordinates sensed by the digitizer cursor which includes a display for depth data, a GO button and a microphone switch. The microphone switch must be forward for the microphone to be live. Included on the cursor module are the standard cursor control buttons as well. Depth data values are inputted by voice, one digit at a time, with a slight pause between digits. The recognized digits appear immediately on the cursor display (but not yet on the self-scan display). Any digit observed to be in error can be deleted by the control word ERASE. A dash (-) will replace the word deleted by ERASE. The operator must replace any dash by a new digit before pressing the GO button. Saying CANCEL erases all dashes. After the operator enters the depth value (one to four digits), he or she should verify the correctness of the complete depth data reading. If the reading is correct, the GO button on the cursor module is depressed firmly, but briefly. During the brief period that the GO button is depressed the position of the cursor on the digitizing table is sensed by the Nova 1200. Any movement before the GO button is released will result in faulty X and Y coordinate data.

After a depth data entry has been made by the use of the GO button, that entry will appear on the self-scan display and as an aid to the operator, it will remain on the cursor display. Also, four small marks will appear to the right of the entered depth value on the cursor display after a GO. If the next depth value is similar to the last, but differs in only one or two digits, the operator need only say the digits which have changed. For example, if an entry is 1 2 3 4 and the next entry is 1 2 3 6, only the least significant digit has changed, 4 to 6. Therefore, after the entry 1 2 3 4 has been entered by voice and the GO button has been pressed the operator will say "6" and the cursor display will indicate 1 2 3 6. The GO button can then be depressed and the next entry will be stored in the active data buffer of the minicomputer memory. If the consecutive entries were 1 2 3 4 and 1 2 5 6, then the second voice entry would be a "5" followed by a "6", etc. The old entry shown on the display can always be deleted by the control word CANCEL. However, after the GO button has been depressed, the control word ERASE is not operable until one or more new digit entries have been made. The current entry will be shown on the right hand part of the self-scan display after the GO button is pressed. Previous entries on the self-scan display move to the left each time GO is pressed.

At any time, any one of up to five previous depth entries may be modified

by the use of the control word, BACKUP. However, only the last three are shown on the self-scan display. Each time this control word is spoken, the cursor display changes to show a depth data value entered just previous to the one currently displayed. For instance, if this command is issued once, the display will not change because it always shows the value inputted before the GO button was pushed. However, it will now be possible to modify the depth value inputted before the GO button was pressed. This can be accomplished by inputting new digits, aided or not by the CANCEL or ERASE control words, as necessary. The BACKUP command can be given up to and including five times. After the appropriate number of BACKUP commands have been uttered and the erroneous data corrected, the original data shown by the cursor display (when the GO button was pressed) can be restored by again pressing the GO button. This time no X and Y data is inputted because GO is merely used to tell the computer that the operator has made the correction necessary and is ready to input new data conditions. If BACKUP is said one, two, or three times and a change is made, the changed depth value will be shown on the self-scan display after GO is pressed. This happens because the last three depth values are shown on the self-scan. Therefore, if any of these are changed the new value will be shown to the operator on the self-scan. If the BACKUP command is issued one or more times but no depth readings are found in error, then no changes should be made by voice, but the GO button must be depressed before inputting new data. After a BACKUP or series of BACKUP commands, care must be exercised to depress the GO button only once before new depth data is entered.

After the first ten depth entries have been made, a card containing the first five entries will be punched. After the next five entries, another card will be punched, etc. If at any time the operator wishes to terminate a data input session, the entries which remain in the computer must be punched on cards by the use of the PUNCH command. After the operators say punch the first time, the self-scan display will show the message "PUNCH?". This tells the operator that the ASR has recognized the word PUNCH. The operator must say either PUNCH again or ERASE. At this time no other word will be accepted. If PUNCH is said the second time, the self-scan will display the message "OK" and the remaining data will be punched on one or two cards. If five or fewer entries remain in the computer buffer only one card will be punched. If six to nine entries remain, two cards will be punched. The self-scan will then display "SYSTEM IS READY" and a new task can be started. Saying ERASE after the first PUNCH resumes normal operation.

Section III

STRUCTURED VOCABULARY WORD RECOGNITION SYSTEM

A. Introduction

To fully satisfy the requirement of the second phase of this contract to fabricate an advanced development model of a highly reliable speaker dependent, limited-vocabulary, word recognition system (WRS), TTI has supplied its commercial limited-vocabulary isolated word recognition system the VIP-100. The VIP-100 version which was supplied for this program includes as principal components the recently redesigned speech preprocessor (part No. 8040) and a Data General Nova 2/10 with 16K of core memory. Interfaced to the computer is a Xebec flexible disk system, model XFD-100, which is used for bulk storage of speaker reference data and node plan data as well as storage of the basic object program should it need to be reloaded at any time. Other speech recognition programs can also be stored on the flexible disk for future use. A Teletype model 33 ASR is used for control and for inputting node plans for the structured vocabulary. The output display module supplied is based upon a 16 character Burroughs self-scan panel. This display panel is of the matrix type of gas discharge display. It has the capability of displaying any of 64 different characters including all digits and alphabetical letters. The supporting software allows display of any word or message up to 16 characters in length upon the recognition of any vocabulary word. Telex 1200 noise cancelling microphones are provided for voice input to the system.

The contract requirement for a structured vocabulary consisting of any combination of nodes utilizing a 64 word vocabulary has been substantially exceeded by the final software configuration. The object program provided with the equipment allows recognition of up to 30 words in any one node of a structured vocabulary which can include up to 30 nodes in the sentence structure. The total vocabulary size is 200 words. Any vocabulary word (except three command words) can appear in any number of nodes in any combination. Two types of operation are possible, sequential and optional. In sequential operation a talker must follow a predetermined sequence of nodes (with one to 30 nodes in the complete sequence) when inputting speech data. Each node can be terminated and the next node in sequence made active by the command word GO. Alternatively, the optional type of operation allows an operator to choose by voice command any of 30 nodes for use at a time. The chosen node is made active by inputting the name of the node by speech. The GO command in this type of operation terminates the current node and allows input of a name of the next node to be used.

B. VIP-100 Operation

The VIP-100 is an adaptive system which can be trained for individual talkers and/or words. Consequently, the system can be automatically adjusted or "tuned" to the voice characteristics of different users in a very short time period. By the inputting of a small number of training samples into the device to provide a reference set of features, the decision criteria for each word in the vocabulary can be modified or trained in an optimum manner.

Thus, the system stores in memory an individual reference set of word features for each word in the vocabulary and for each talker in the system. Once system training is completed, new words spoken into the device during normal operation are compared with the stored references and a "closest fit" is selected as the recognized word. It is also possible to obtain a "no decision" or reject, when the characteristics of several words in the reference memory are very close to the spoken word. Since rejects may be permitted a predetermined percent of the time, a trade-off can be made between a reject (the speaker must repeat the word) and possible false responses. With this trade-off, it is possible to achieve high recognition accuracies and small substitution errors. The decision technique employed can be described most simply by briefly reviewing the operation of the system training and recognition mode.

During the training mode, the VIP-100 automatically extracts a time-normalized feature matrix for each repetition of a given word. A consistent matrix of feature occurrences (between repetitions) is required before the features are stored in the reference pattern memory. A template threshold factor has been chosen such that a feature occurrence (in a given time segment) is considered valid only when it occurs a minimum number of times relative to the number of training samples. In the operational mode, each new word spoken into the system is processed in a manner analogus to the training procedure--i.e., features extracted, digitized and time normalized. The resultant test word matrix then is compared digitally to each stored reference matrix. Similarities and dissimilarities in each compared matrix are appropriately weighted and the net result provides a weighted correlation product. Correlation products also are generated after shifting the input word matrix ±1 time segment. The stored reference word producing the highest overall correlation is selected as the test word.

C. System Configuration

Figure 4 is a block diagram of the complete speech recognition system. All units except the ASR 33 Teletype, the self-scan display and the Voice Input-Remote Control are mounted in a 72 inch high rack which is complete with cooling fans. The Voice Input-Remote Control unit can be located at the convenience of the operator and is used to amplify the input speech and to adjust system audio gain to a convenient level for the operator. This unit also houses thumbwheel switches which can be set to designate a word number for training purposes. If the operator desires to train or retrain a word or the entire vocabulary he or she can select the appropriate number and press the train button located on this unit. The speech preprocessor/recognition processor accepts this new training data and processes it such that the appropriate word reference data is stored in the minicomputer memory in place of the existing data for the word trained.

D. System Software

The computer program supplied with the WRS has a total vocabulary capability of 200 words. The words 10 through 42 are restricted. The remainder of the vocabulary may be selected by the user. Up to and including 30 nodes may be included in the structure with up to and including 30 words in each

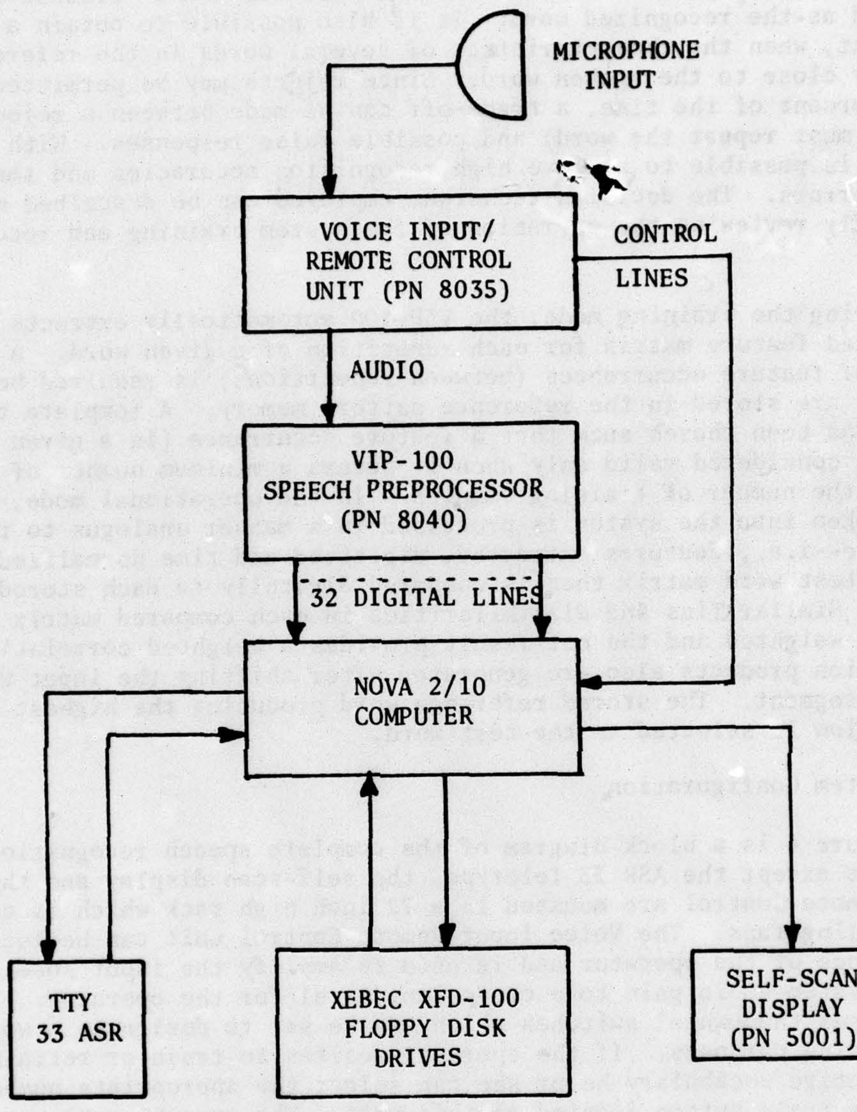

Figure 4. Block diagram of the Advanced Development Model isolated-word recognition system.

node. A particular vocabulary word may be included in any number of nodes. Every vocabulary word has a number from 0 thru 199 which must be used for constructing nodes, for training and for constructing display messages. Any vocabulary word (except 10, 11 and 12) may be represented on the output display by up to 16 letters (or digits or combination thereof). Vocabulary word 10 is the command word GO, 11 is the command ERASE and word 12 is the command CANCEL. All other words are chosen by the operator. The command GO is always active. It services to terminate the active node and to make active the next node in sequential or allow input of a new node name in optional operation. The command ERASE is always active and erases the last spoken word. In sequential operation, the node number and name cannot be erased. The command CANCEL is always active but has slightly different functions in optional and in sequential operation. In optional operation CANCEL deletes all spoken inputs in the active node as well as the node itself. In sequential operation, CANCEL deletes all spoken inputs in a node but cannot delete the node.

Each node must be identified for display by a node name and for recognition by the number of the vocabulary word which will call that node in the optional type of operation. Only vocabulary words 13 through 42 can be used to call new nodes. Therefore, node numbers must be 13 through 42. Node names for display can be the same as the vocabulary words used to call them. This is NOT mandatory, however. The node names are assigned when constructing the node plan. Vocabulary word names are assigned independently. Node names appear on the display in optional operation when a node is called by a vocabulary word from 13 to 42. Vocabulary word names appear on the display when inputting data or during training. In sequential operation, node names on the display appear after a GO to indicate the next active node. The order of nodes in this latter operation is from lower to higher node numbers. When the program starts in the sequential mode, the lowest numbered node is always active first. After all nodes have been completed, the program reverts again to the lowest numbered node.

Flowcharts of the software supplied are shown in Figure 5. The first sheet of the set of three flowcharts illustrates the sequential type of operation. When sequential operation is selected by Teletype commands, as shown, the lowest numbered node in the current plan is made active. That is, only the vocabulary words selected for that node are made eligible for recognition. Also, the name of the node is shown on the display. The operator can then input by voice any words in that active node or go to the next node by saying GO.

The second and third flowchart sheets show optional operation. Optional operation is also selected by keyboard command as shown by flowchart sheet one. When this type of operation is selected, no node is initially made active. Instead, vocabulary words 13 through 42 are eligible for recognition. These words are used to call the nodes, i.e. make them active. Initially, the display will show the message "NAME THIS NODE". At any time the command GO will deactivate a node and again make words 13 through 42 eligible for recognition so that another node may be called.

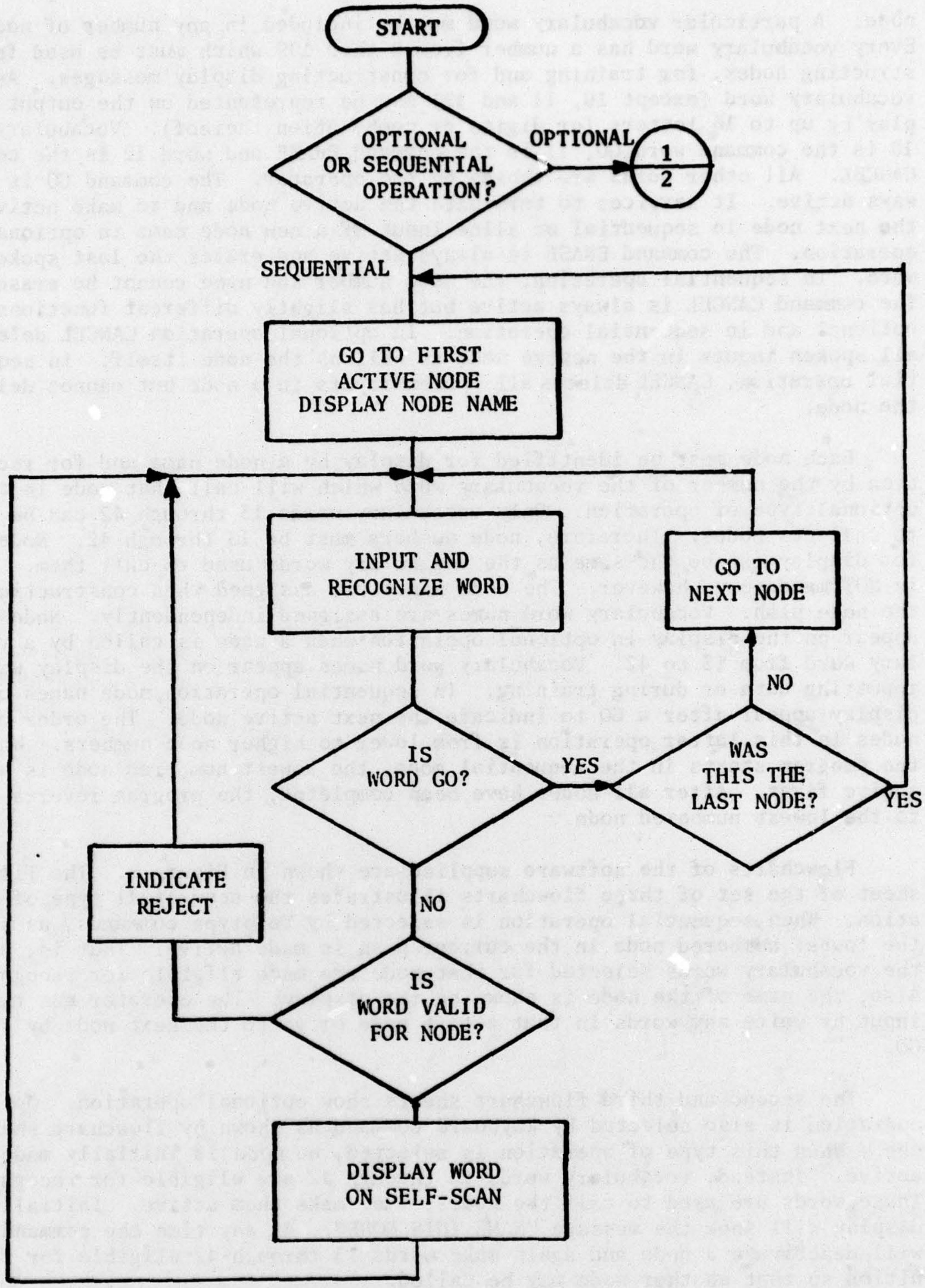

Figure 5. Flowchart of 200 word structured program (Sheet 1 of 3)

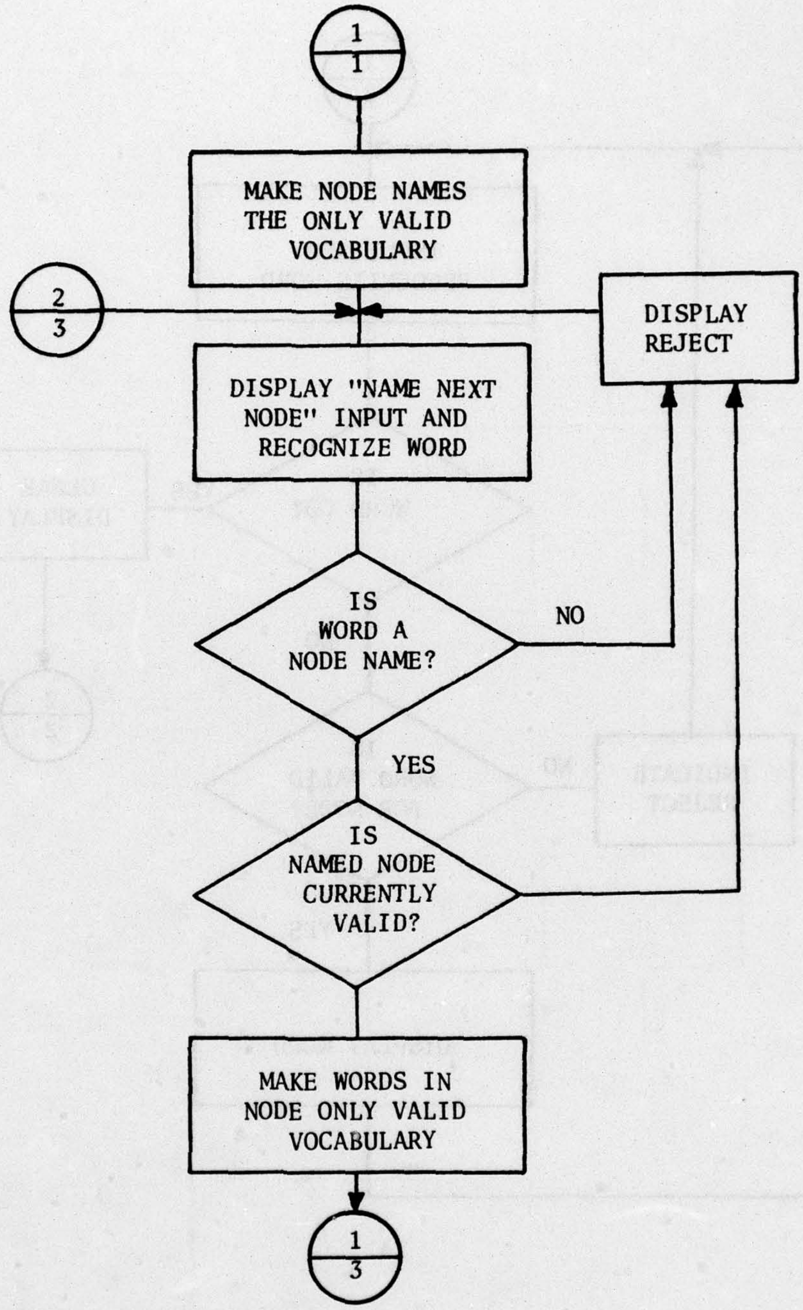

Figure 5, Flowchart of 200 word structured program (Sheet 2 of 3)

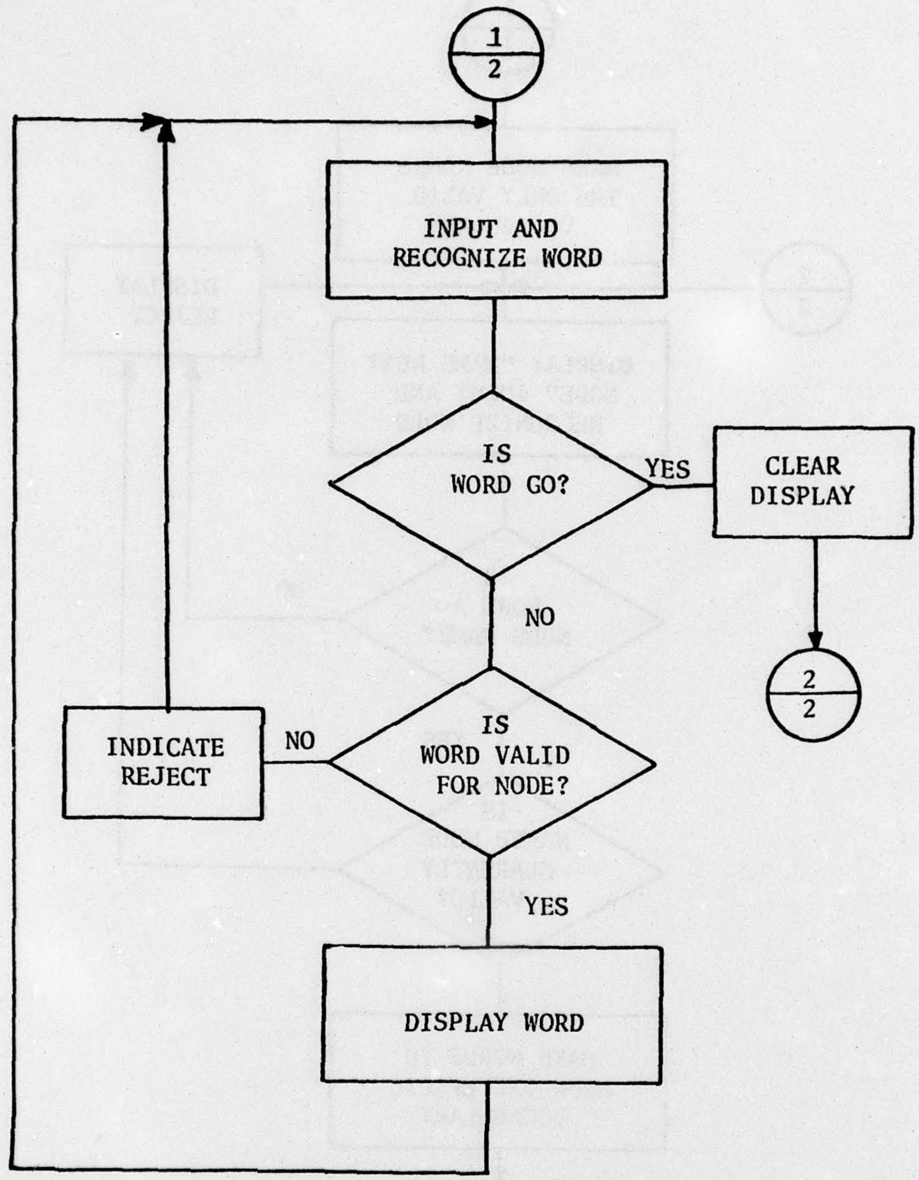

Figure 5. Flowchart of 200 word structured program (Sheet 3 of 3)

E. System Control By Teletype Commands

Before any words are spoken or any Teletype commands are issued the system must be properly started. The computer, the speech preprocessor, the display, the floppy disk drive and the Teletype all have separate on-off switches. If all are connected to the AC supply outlets in the rack, all power can be switched by a single system power switch. After power is on to all components, pressing the Reset and the Start switch causes the following message on the TTY:

TYPE 1 FOR INSTRUCTIONS

Typing a 1 will result in the following Teletype message:

STRUCTURED RECOGNITION PROGRAM FOR 2ØØ WORDS

TYPE:

T - ASSIGN TRAINING PARAMETERS
I - INPUT TRAINING DATA
O - OUTPUT TRAINING DATA
A - INPUT NODE DATA
B - OUTPUT NODE DATA
G - GO TO RECOGNITION PHASE
S - START THE SYSTEM
M - MODIFY DISPLAY MESSAGES FOR WORDS IN VOCABULARY
H - PRINT ACTIVE NODES TOGETHER WITH WORDS IN EACH ACTIVE NODE
V - PRINT VOCABULARY
Q - EDIT NODE STRUCTURE
C - MODIFY NODE DISPLAY MESSAGES

The above commands and the operations which they each activate are explained in the following paragraphs.

1. Command T - For Assigning Training Parameters

The system has a capacity of 200 words. If a smaller vocabulary is to be used for a particular application the vocabulary size should be set to the desired size before training by the use of this keyboard command. Typing the letter "T" (remember that all keyboard entries must be terminated with a carriage return) will cause the Teletype to replay "NO. OF REPS?". The user should replay with a number from 1 to 10 to indicate the number of repetitions of each vocabulary word which will be spoken to train the system. The most reliable recognition results are obtained when 10 training samples are used. Next the Teletype will ask for the vocabulary size by responding, "VOCABULARY SIZE?". The size from 1 to 200 should then be typed.

Actual training (or retraining) can be accomplished at any time by the use of the OPERATOR and WORD thumbwheel switches to indicate the number of the word to be trained and the use of the TRAIN button to initiate the training process. These switches and button are located on the Voice Input-Remote Control Unit. Setting the switches at the vocabulary size inputted by Teletype as outlined above will cause the entire vocabulary (of that size) to be trained. Training can be halted at any time by setting the thumbwheel

switches to a number larger than the vocabulary size and pushing TRAIN button.

During training, the operator for whom the system is to be trained should properly position the microphone and set the volume to the appropriate level. When the complete vocabulary is trained, consecutive samples of a given vocabulary word are entered in sequence. That is, all samples of the first vocabulary word should be entered first. The display will then indicate the second word to be trained and continue displaying that word until all training samples have been entered. The process will be continued until the entire vocabulary is trained.

2. Commands S and G - For Inputting Speech Data

The system can be put into one of two speech recognition states by the use of the command S - to start the system. The Teletype will type one of the following two messages (depending on past operation):

(1) SYSTEM IS SET FOR SEQUENTIAL NODES, IF OK TYPE 1; or
(2) SYSTEM IS SET FOR OPTIONAL NODES, IF OK TYPE 1

The operator can either accept the type of operation indicated by typing 1 and a CR or can go to the other choice by typing any other character and CR. The command G will go directly to recognition without the option of changing the manner of operation.

3. Command Q - For Editing Nodes

This allows a simple editing function for the nodes and the words associated with each node. The message:

A,D,T?

is typed and the user types one of the above to select a function.

A - Add a node
D - Delete a node
T - Terminate the edit

Note - to change the vocabulary of a node, the node is first deleted then added with the appropriate changes.

a. The ADD Function

Typing an A causes the TTY to respond:

"NODE #"

The valid response is a number 13-42. If a NODE exists with that number, the program repeats its request. Every NODE must have a number. That number is the same as the number of the vocabulary word which calls that node in optional operation.

"NAME OF THIS NODE" - input is up to 16 alphanumeric characters and is the message which will appear on the display when this NODE is activated in either sequential or optional operation. No more than 16 characters will be accepted.

"WORD SET SAME AS NODE #?" - allows this node to have the same conditions as an existing node. Typing a valid node number completes the ADD function, and the TTY responds with "A,D,T?". Typing a non-existent node number causes the program to advance.

"WORD #" - typing any number which corresponds to a vocabulary word attaches that word as a condition to the category. To terminate this process the user types 200 and the program requests "A,D,T?". Message which may appear during the ADD function:

"NODE LIST FILLED"

There is no room for the node. The program can accommodate up to 30 nodes. NOTE - It is extremely important that only a well ordered exit be made from the ADD function. If an error is made during ADD, the function must be completed.

b. The DELETE Function

Typing a D causes the TTY to respond "NODE #" - the number of the category to be deleted. An illegal number causes the request to be repeated. It is also possible to make a well ordered exit from the Edit function at this point by typing CTRL-P. Message which may appear during DELETE:

"ALL NODES DELETED"

c. The TERMINATE Function

Typing a "T" causes the editing of nodes to be terminated and the TTY will type the message "TYPE 1 FOR INSTRUCTIONS" after which another command can be issued.

4. Command C - For Modifying Node Names Without Changing Node Structures

The TTY responds to a "C" by outputting the message "NODE #" - the number of the node whose message is to be changed. Typing a non-existent category number causes this function to terminate. Typing a valid number causes a line feed and the program waits for input of up to 16 characters followed by a carriage return.

5. Command B - For Outputting Node Data to Floppy Disk

The node structure entered by the use of the command Q may be saved for future use, on floppy disk. The operator must first mount an initialized floppy disk on FD-1 (right-hand drive). Then he types B with a CR. The Teletype will respond with a message which was stored on the disk (if any). Next he types the plan number, (decimal from 1 to 20) followed by CR. Each disk holds 20 node plans. Display vocabulary names are also stored on disk together with node plans.

6. Command A - For Inputting a Node Structure

A node structure and reference which have previously been outputted to disk on tape as outlined above may be inputted when needed. Again the node plan disk is placed on FD-1 and the letter A is entered by Teletype. The Teletype will ask for plan number (from 1 to 20, decimal). If a non-existent number is typed bad data will be inputted.

7. Command O - For Outputting to Floppy Disk of Training Data

The reference data compiled during training may be saved on floppy disk for future use. The resulting disk data will retrain the system for the particular operator and vocabulary when it is read into the system with the appropriate command. The reference data is produced with the output "O" command after an initialized disk has been mounted on FD-1. The Teletype will respond with a message which was put onto the disk previously (if any) and ask for speaker number which can be any decimal number (1 to 20) followed by a CR. Node plan data and speaker data cannot be written on the same disk.

8. Command I - For Inputting Training Data from Floppy Disk

The system may be trained from a previously produced reference data which has been stored on a floppy disk. The floppy disk holding the speaker data is mounted on FD-1. The I command is typed and the Teletype then requests the speaker number. The appropriate speaker number (1 to 20 decimal) is typed followed by a CR. If that speaker is not on the disk, random data will replace the valid training data in the computer. Therefore, it is very important to keep accurate records of speaker numbers (and node plan numbers as well).

9. Command M - To Modify Display Messages for Vocabulary Words (Regardless of Node Structure)

Two message modify entry techniques are available if the operator does not wish to retain any of the previously displayed message. He may enter a full 16 characters during the message modify instruction by using spaces to fill character positions not needed for the actual message. As an alternative, he can enter control A (hold down control key while pressing the A key) as the first character of the message and then enter the message he wants displayed. The control A character will cause the display to be cleared of all previously displayed characters before the new message is displayed. The operator must enter the correct number of leading spaces with either technique if the message is to be centered in the display.

The message modify routine is called by typing M on the Teletype. The Teletype will respond with "WORD NO.?". The operator should reply with the number of the first word for which the display message is to be modified (remember that the first word is word number Ø). The Teletype will respond with a carriage return and line feed. The operator should then enter the new message to be displayed for the particular vocabulary word. A total of 16 or less characters including spaces and control characters should be entered; the message should be terminated with a carriage return. The Teletype will

then ask for the word number of the next word for which the display message is to be modified. The rub-out feature is not operational during the character entry procedure; if pressed, it will appear as a "?" in the displayed message.

Three special control characters are available during the character entry procedure. They are called:

Control A--clears the display of current message
Control B--blanks the display for approximately 250 milliseconds
Control C--backspaces the current message one position

Control may be returned to the selection routine when all desired message modifications have been completed. This is accomplished by answering the "WORD NO.?" request with 200.

10. Command H - For Printing Node Vocabulary

Typing the command H will result in a Teletype printout of each node vocabulary list. The nodes will appear in numerical order. Each node list will be headed by the node number (to the left) and the node name (to the right). Below the node number will be numbers of the words in the node vocabulary. Below the node name will be the display messages for the node vocabulary words.

Section IV

CONCLUSIONS AND RECOMMENDATIONS

The major objectives of the Alpha/Numeric Extraction Technique Phase II program have been achieved. The word recognition system (WRS) developed under the predecessor to this program (Contract F30602-75-C-0238) has been interfaced to a bathymetric digitizing system at the Defense Mapping Agency Hydrographic Center (DMAHC), Suitland, Maryland. The incorporation of voice input into the bathymetric digitizer system should allow more rapid and accurate entry of depth data from smooth sheets. The replacement of the digitizer keyboard with voice data entry allows the operator to concentrate on accurate location of the cursor and accurate entry with verification of the depth sounding data. The operator need not divert his or her attention from the cursor to the keyboard after noting each depth reading in order to input depth data. Rather, the operator can input the depth data by voice while accurately aligning the cursor and then quickly shift to the next data point.

The advanced development model of a word recognition system (WRS) with a 200 word structured vocabulary which was also developed under this program should have wide potential in the cartographic area. This system can be adapted to a number of specific tasks by the use of different node plans for the structured vocabulary. Node plans as well as speaker reference data and vocabulary display messages can be stored on floppy disks with the disk drive included in the system. Node plan data and speaker reference data can be retrieved in seconds from a floppy disk.

In order to determine the effectiveness of the DMAHC voice-digitizer system a series of tests should be conducted in which manual depth input is compared with voice depth input. Such tests should be conducted over a period of time so that a learning curve can be observed. Such a curve would show how operators adapt to the use of voice for depth input and how their speed and accuracy increases as they adapt. Speed and accuracy benchmarks for manual and voice entry should be recorded during such tests.

Reference

1. P.B. Scott, "Alpha/Numeric Extraction Technique", Tech. Report RADC-TR-75-287., (B008303) dated Nov 75.

METRIC SYSTEM

BASE UNITS:

Quantity	Unit	SI Symbol	Formula
length	metre	m	...
mass	kilogram	kg	...
time	second	s	...
electric current	ampere	A	...
thermodynamic temperature	kelvin	K	...
amount of substance	mole	mol	...
luminous intensity	candela	cd	...

SUPPLEMENTARY UNITS:

plane angle	radian	rad	...
solid angle	steradian	sr	...

DERIVED UNITS:

Acceleration	metre per second squared	...	m/s
activity (of a radioactive source)	disintegration per second	...	(disintegration)/s
angular acceleration	radian per second squared	...	rad/s
angular velocity	radian per second	...	rad/s
area	square metre	...	m
density	kilogram per cubic metre	...	kg/m
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per metre	...	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm		V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	...	J/K
force	newton	N	kg·m/s
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m
luminance	candela per square metre	...	cd/m
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per metre	...	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m
magnetomotive force	ampere	A	...
power	watt	W	J/s
pressure	pascal	Pa	N/m
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	...	W/sr
specific heat	joule per kilogram-kelvin	...	J/kg·K
stress	pascal	Pa	N/m
thermal conductivity	watt per metre-kelvin	...	W/m·K
velocity	metre per second	...	m/s
viscosity, dynamic	pascal-second	...	Pa·s
viscosity, kinematic	square metre per second	...	m/s
voltage	volt	V	W/A
volume	cubic metre	...	m
wavenumber	reciprocal metre	...	(wave)/m
work	joule	J	N·m

SI PREFIXES:

Multiplication Factors	Prefix	SI Symbol
1 000 000 000 000 = 10^{12}	tera	T
1 000 000 000 = 10^{9}	giga	G
1 000 000 = 10^{6}	mega	M
1 000 = 10^{3}	kilo	k
100 = 10^{2}	hecto*	h
10 = 10^{1}	deka*	da
0.1 = 10^{-1}	deci*	d
0.01 = 10^{-2}	centi*	c
0.001 = 10^{-3}	milli	m
0.000 001 = 10^{-6}	micro	μ
0.000 000 001 = 10^{-9}	nano	n
0.000 000 000 001 = 10^{-12}	pico	p
0.000 000 000 000 001 = 10^{-15}	femto	f
0.000 000 000 000 000 001 = 10^{-18}	atto	a

* To be avoided where possible.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.